

Do Global Models Properly Represent the Feedback between Land and Atmosphere?

PAUL A. DIRMEYER

Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland

RANDAL D. KOSTER

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

ZHICHANG GUO

Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland

(Manuscript received 10 August 2005, in final form 5 January 2006)

ABSTRACT

The Global Energy and Water Cycle Experiment/Climate Variability and Predictability (GEWEX/CLIVAR) Global Land–Atmosphere Coupling Experiment (GLACE) has provided an estimate of the global distribution of land–atmosphere coupling strength during boreal summer based on the results from a dozen weather and climate models. However, there is a great deal of variation among models, attributable to a range of sensitivities in the simulation of both the terrestrial and atmospheric branches of the hydrologic cycle. It remains an open question whether any of the models, or the multimodel estimate, reflects the actual pattern and strength of land–atmosphere coupling in the earth’s hydrologic cycle. The authors attempt to diagnose this by examining the local covariability of key atmospheric and land surface variables both in models and in those few locations where comparable, relatively complete, long-term measurements exist. Most models do not encompass well the observed relationships between surface and atmospheric state variables and fluxes, suggesting that these models do not represent land–atmosphere coupling correctly. Specifically, there is evidence that systematic biases in near-surface temperature and humidity among all models may contribute to incorrect surface flux sensitivities. However, the multimodel mean generally validates better than most or all of the individual models. Regional precipitation behavior (lagged autocorrelation and predisposition toward maintenance of extremes) between models and observations is also compared. Again a great deal of variation is found among the participating models, but remarkably accurate behavior of the multimodel mean.

1. Introduction

If robust interactions between the slowly varying land surface state and the atmosphere on weather-to-climate time scales could be demonstrated, predictions of the climate system could be enhanced. Monitoring of land surface states to initialize numerical models with the proper coupling between terrestrial and atmospheric processes should lead to improved forecasts. There are many global and regional modeling studies that suggest such interactions exist, especially involving the land surface state variable of soil wetness. But these modeling results have largely been based on long simu-

lations, ensemble simulations, or large area averages that outstrip the coverage of current observational datasets. Therefore, observational evidence to back up the finding of the models is scarce. In addition, different models have shown different character or degrees of response, casting an additional shadow of uncertainty over the prospect of exploiting land–atmosphere interactions for enhanced predictability. Even in analytical models with highly controlled parameters, varied and conflicting results arise (e.g., Findell and Eltahir 2003a,b; D’Odorico and Porporato 2004; Kochendorfer and Ramirez 2005; Teuling et al. 2005). Two questions arise. Is there a model consensus regarding land–atmosphere feedbacks? Are any models (or is the consensus) close to being correct?

Recently, an international initiative was undertaken by a dozen weather and climate modeling groups including both operational and research centers to deter-

Corresponding author address: Dr. Paul A. Dirmeyer, Center for Ocean–Land–Atmosphere, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705-3016.
E-mail: dirmeyer@cola.iges.org

mine the degree to which the atmospheric branch of the hydrologic cycle is coupled to the land surface within global coupled land–atmosphere models. The project, called the Global Land–Atmosphere Coupling Experiment (GLACE), is jointly sponsored by the Global Energy and Water Cycle Experiment (GEWEX) Global Land–Atmosphere System Study (GLASS) and the Climate Variability and Predictability (CLIVAR) Working Group on Seasonal–Interannual Prediction (WGSIP). Participating GCMs include those from the Bureau of Meteorology Research Centre (BMRC) and Commonwealth Scientific and Industrial Research Organisation Conformal-Cubic 3 (CSIRO-CC3) in Australia, the Canadian Climate Centre (CCCma), the Center for Climate System Research (CCSR) at the University of Tokyo, the Hadley Centre in the United Kingdom (HadAM3), and seven from the following centers in the United States: the Center for Ocean–Land–Atmosphere Studies (COLA), the Geophysical Fluid Dynamics Laboratory (GFDL), the National Center for Atmospheric Research [NCAR; Community Atmosphere Model 3 (CAM3)], the National Centers for Environmental Prediction [NCEP; Global Forecast System/Ohio State University (GFS/OSU)], the University of California at Los Angeles (UCLA), and two from the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center [Geostationary Earth–Orbiting Satellite–Climate Radiation Branch (GEOS-CRB) and NASA Seasonal-To-Interannual Prediction Project (NSIPP)].

As described in detail by Koster et al. (2006), each modeling group in GLACE was asked to perform an ensemble of 16 three-month simulations with its general circulation model (GCM) beginning on 1 June and using the same specified sea surface temperature for all simulations (case W). The ensemble members vary only in their initialization, preferably taken from 1 June states of a multidecade, parallel integration, so that the members would be as independent as possible. Each group chose one member to be the basis of test case ensembles, and saved all land surface state variables at every model time step from that member. Two test ensembles were made—one with all land surface state variables specified (i.e., prescribed at each time step) to match the chosen member from the control ensemble (case R), and the other having only soil wetness specified for soil layers below the thin surface layer (case S). Comparisons between the test ensembles and the control ensemble show to what degree elements of the land surface affect seasonal climate.

Koster et al. (2004, 2006, hereafter referred to as K04 and K06, respectively) showed the global distribution

of the strength of land–atmosphere feedback, manifested in precipitation, as calculated across the 12 models. “Hot spots” appeared for boreal summer over several parts of the world, including the Great Plains of North America, sub-Saharan Africa north of the equator, India, and parts of China. The signal was generally weak over the Southern Hemisphere (austral winter), high latitudes, and very arid or humid regions. K06 also showed a great deal of variation among models, both in terms of patterns and the overall strength of feedbacks. The multimodel pattern of hot spots is not plainly evident in several of the individual models.

Guo et al. (2006, hereafter G06) showed that the pathway for strong feedbacks in the models requires both a robust coupling of surface fluxes to soil wetness in the land surface component of the model, and a strong link between precipitation and surface fluxes in the atmospheric model through convection. G06 was able to separately quantify these two segments of the feedback loop in the models and show that weakness in either branch hindered the overall link between soil wetness and precipitation. Furthermore, the land surface segment was found generally to be weak in humid regions (due to a lack of evapotranspiration response to changes in soil wetness) and arid regions (because of low variability), while the atmospheric segment was weak in arid zones. This leaves the transitional regions between arid and humid as the only regions where both segments can propagate information about soil wetness anomalies to the convective parameterizations and exert some control on precipitation.

K06 and G06 use the metric of “coupling strength,” denoted by the symbol Ω , in the multimodel assessment of land–atmosphere feedbacks; Ω is a measure of the coherence of a seasonal time series of a prognostic or diagnostic model variable (e.g., precipitation or evaporation) across a range of ensemble members that have been initialized differently. In essence, it is the fraction of the total variance of a variable that is “explained,” or forced, by the prescription of all boundary conditions in the model. The coupling strength between land and atmosphere is quantified by the change in Ω ($\Delta\Omega$) between an ensemble with differently initialized, freely evolving land surface state variables, and an ensemble where the land surface state variables (namely, subsurface soil wetness) are specified to match one case from the control ensemble. The idea is that if the land surface is exerting some controlling influence on surface fluxes and atmospheric processes, the restriction in the time evolution of the land surface state variables should result in an increase in the coherence among the time series of surface fluxes and meteorological states. Feedbacks are implied by a nonzero value of $\Delta\Omega$ (which

varies nominally from 0 to 1), with the degree of coupling measured by the magnitude of $\Delta\Omega$.

The parameter Ω is a handy construct for model comparisons and analysis, but it is not a physical quantity. It is an artifact of ensemble simulations. The real world does not present us with an ensemble of parallel histories, but only one realization. Therefore, there is no direct way to calculate a field of Ω , much less $\Delta\Omega$, from observations. This is but one of the impediments to quantifying the land–atmosphere coupling strength in the environment.

Another difficulty is the lack of global measurements of soil moisture and surface fluxes, which are key elements of the coupling pathway. There have been efforts to infer feedbacks from the observational record. Betts et al. (1996) show from field data collected at middle and high latitudes that the interaction of the land surface and the atmosphere is primarily through its influence on the character of the planetary boundary layer (its depth, moisture content, rate of entrainment of air from above, and its ability to trigger convection) as a result of surface properties such as soil wetness, vegetation, and the diffusivity of heat in the soil column. Findell and Eltahir (1997) showed a positive correlation between variations in the observed soil moisture records in the Illinois Climate Network and rainfall in the subsequent three weeks, which they claimed was observational evidence of a positive hydrologic feedback between land and atmosphere. Salvucci et al. (2002) subsequently showed that the formulation of the calculation by Findell and Eltahir (1997) biased the results by allowing some future soil wetness information to affect the correlation. Koster and Suarez (2004) showed that there is a statistically significant separation in the probability density functions of monthly rainfall during summer over midlatitude land, conditioned on the rainfall anomaly during the previous month. Parallel analyses of GCM simulations suggest that other factors (e.g., alteration of the circulation due to remote SST anomalies) are not responsible for the separation, and thus the separation implies a positive feedback between land and atmosphere.

We have in the GLACE results, a multimodel-based estimate of the strength and spatial variation of land–atmosphere coupling, and its relationship to state variables and fluxes within global models. Can we confirm or refute the GLACE results using the observational record? Where thorough surface flux and land state observations exist, we attempt to validate the GLACE models and to establish relationships among measured and unmeasured (purely model-derived) variables that may allow us to infer more about the veracity of the GLACE results. The recent paper by Betts (2004; here-

after B04) provides a framework, based on a series of relationships found in an independent global model, which can be followed to quantify underlying elements of land–atmosphere coupling strength from measurable quantities at the surface and in the boundary layer.

Section 2 describes the observational datasets that we used. In section 3, we attempt to link the models' Ω parameter for evapotranspiration to observable quantities and validate the performance of the model simulations. The model validation to other relationships is expanded in section 4. In section 5 we include the atmospheric segment of coupling by comparing the models' behavior to observational evidence of persistence in precipitation anomalies. Conclusions are given in section 6.

2. Observational data

To compare the model representation of land–atmosphere coupling strength to that in the real world, complete observations of land surface state variables, near-surface atmospheric states, and fluxes between land and atmosphere are needed. These observations must span a long enough period of time to provide a large sample covering the range of variability of these variables and to provide adequate statistical significance for the results. Finally, we are interested in the same season as the GLACE experiments, spanning June, July, and August. There are very few sources of observational data that can meet all these requirements. Two are identified for this study.

The U.S. Department of Energy operates the Atmospheric Radiation Measurement (ARM) Program (Ackerman and Stokes 2003). In particular, the southern Great Plains site consists of a Central Facility and a number of Extended Facilities across a large area of Oklahoma and southern Kansas, each having instrument clusters to measure radiation, near-surface meteorology, surface fluxes, soil moisture, and temperature. For our application, data from the Energy Balance Bowen Ratio (EBBR; Cook 2005) system is appropriate. The EBBR is a ground-based sensor system installed over grass that uses observations of net radiation, soil heat flow (25 mm below the surface), surface soil moisture (top 50 mm layer), and the vertical gradients of temperature and relative humidity to estimate the vertical heat fluxes at the local surface by a Bowen ratio energy balance technique. The complete set of near-surface meteorological variables is measured as well. Data archives exist for 14 stations.

We use the B1-level 30-min-average data and average them further to daily time scales for consistency with the model output from GLACE. The summer data

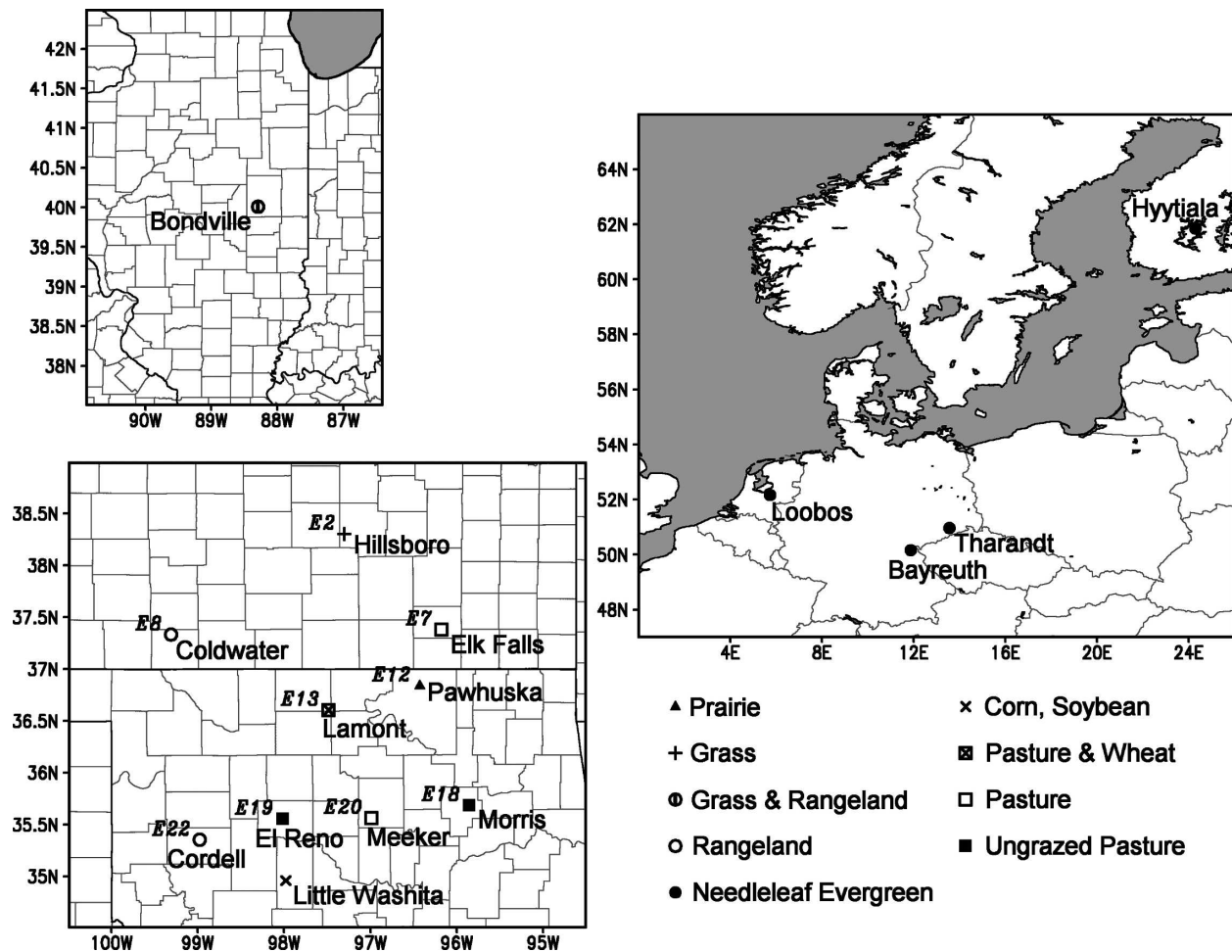


FIG. 1. Location of ARM Extended Facilities (lower left beside station codes starting with "E") and FLUXNET sites used in this study. Symbols indicate the type of vegetation cover.

for the years 2001–04 are used. Applying a rather strict quality filter to the data, we reject any day's data for a variable if there is not at least 21 h of data with no quality control issues flagged. Then stations that have excessive missing data are screened out. We have two criteria; one is that the station must have at least 75% of the days with all terms of the surface heat fluxes (latent, sensible, and ground heat fluxes) available. The second is that 75% of the days must have soil moisture measurements. These criteria eliminate five stations from consideration. Earlsboro (E27) only came online in late 2003, and has intermittent measurements in the archive. Plevna (E4) has intermittent data throughout the 4-yr period. Cement (E26) is missing significant amounts of data during 2001 and 2002. Ashton (E9) has no flux data for 2001 and about half of 2002. Ringwood (E15) is missing most of the soil wetness measurements for 2002–04. This leaves nine stations with sufficient data for comparison with the models. These stations are

shown with their station codes in the lower-left panel of Fig. 1, along with land-cover type (indicated by the symbols). The station data are examined individually and combined to represent averages over scales similar to a GCM grid box.

The second source of data comes from the FLUXNET network of micrometeorological tower sites (Baldocchi et al. 2001). Though designed primarily to measure the exchanges of carbon dioxide, water vapor, and energy between the biosphere and atmosphere, they also include standard meteorological measurements, and in some cases subsurface water and temperature data. In a quest for data that are quality controlled, we have drawn upon the long-term archive at the Oak Ridge National Laboratory Distributed Active Archive Center. Daily gap-filled data (Falge et al. 2003) from the AmeriFlux and EUROFLUX regional networks are available for a number of years. Two AmeriFlux sites [Bondville, Illinois (Meyers and Hollinger 2004), and

Little Washita, Oklahoma (Meyers 2001)] have multi-year records of fluxes and soil moisture (averaged over top 100 mm). None of the EUROFLUX sites in the gap-filled dataset record soil moisture, but four sites (Bayreuth, Tharandt, Loobos, and Hyytiala) have 12 or more summer months of other relevant observations in the archive. Data used from the EUROFLUX sites cover the summers of 1996–2000 except for Bayreuth, which covers 1996–99. The locations and vegetation cover of these sites are also given in Fig. 1.

In addition to sample size and the list of variables, the data must also represent a reasonably closed surface water and energy balance in order to be useful for model validation. Figures 2 and 3 show the surface energy balances (measured net radiation versus the sum of measured surface latent, sensible, and ground heat fluxes) for the ARM and FLUXNET sites, respectively. The bold line is the least squares linear regression of the surface heat fluxes on the net radiation. Also shown are the RMSE, explained variance (r^2), and bias with respect to the perfect fit line. The ARM sites show very tight closure in most cases, with a tendency toward slight positive biases (heat fluxes exceed net radiation). These are Bowen ratio stations, so the good fit is not surprising. As might be expected, the average across the ARM sites has the highest r^2 and the lowest of RMSE. The fit for the FLUXNET sites is not as good, with a tendency for negative biases and greater scatter. Only Tharandt has a bias as low as the ARM sites. Days with greater than 50% gap filling in surface flux terms or radiation are not included in Fig. 3 or the calculations. Note that at the Hyytiala site there are no ground heat flux measurements, so the terms of the surface energy balance are not completely specified, possibly contributing to the appearance of a strong negative bias there. At the other stations in Fig. 3, removal of the ground heat flux term would contribute an average additional bias of -8.0 W m^{-2} to the relationships.

It is worth reminding the reader that the ARM facilities and the Little Washita AmeriFlux site lie near the center of the North American “hot spot” for land–atmosphere coupling identified by GLACE (K04). Thus, we begin with an expectation that the observations from these sites may provide the strongest available evidence for land surface feedbacks on weather and climate. However, there is also evidence that GCMs often do a poor job of simulating climate in this area (e.g., Fennessy and Xue 1997; Koster and Suarez 2003; Sud et al. 2003; GFDL Global Atmospheric Model Development Team 2004). The European sites are in a more quiescent region for land–atmosphere coupling, according to the GLACE models, providing

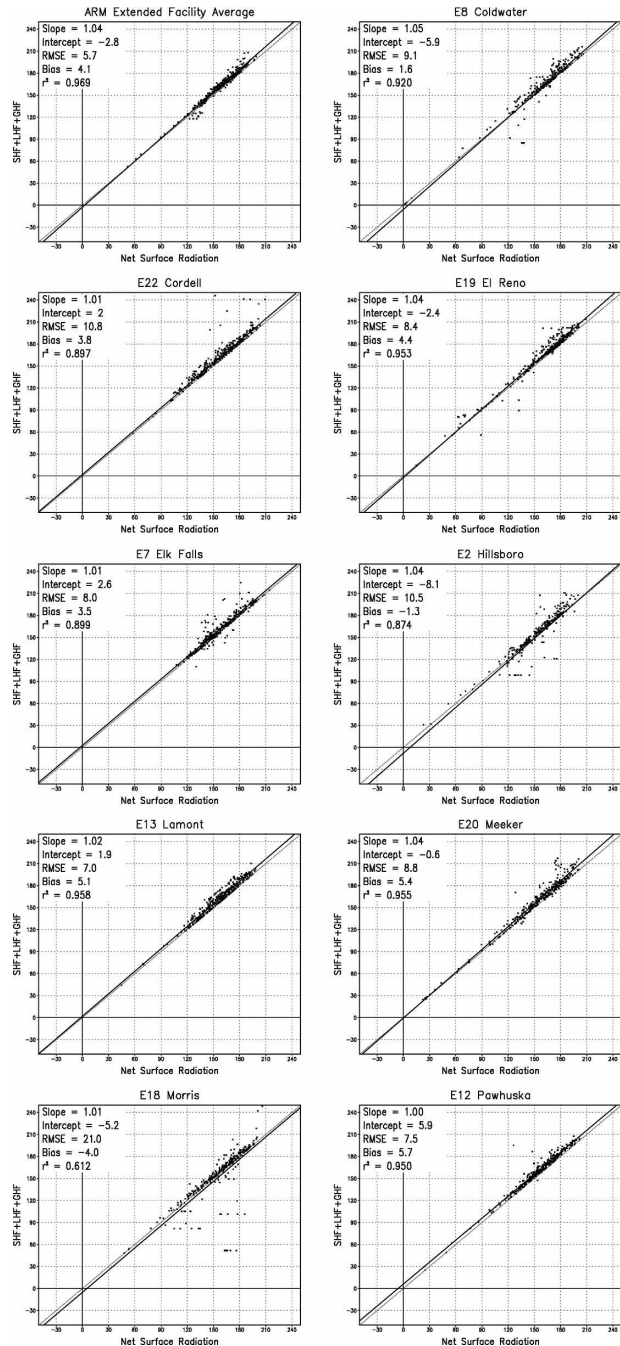


FIG. 2. Validation of the energy balance from 6-day means at selected ARM Extended Facility sites for June–August 2001–04, and the average across all sites (upper left). Units: W m^{-2} . The diagonal dashed gray line shows exact balance; the black solid line is the best-fit linear regression through the data points.

an opportunity to compare and contrast among the models and observations.

All of the scatter diagrams and other comparisons use 6-day averages as in K06 and G06, similar to B04. The analysis in section 5 utilizes two observed precipi-

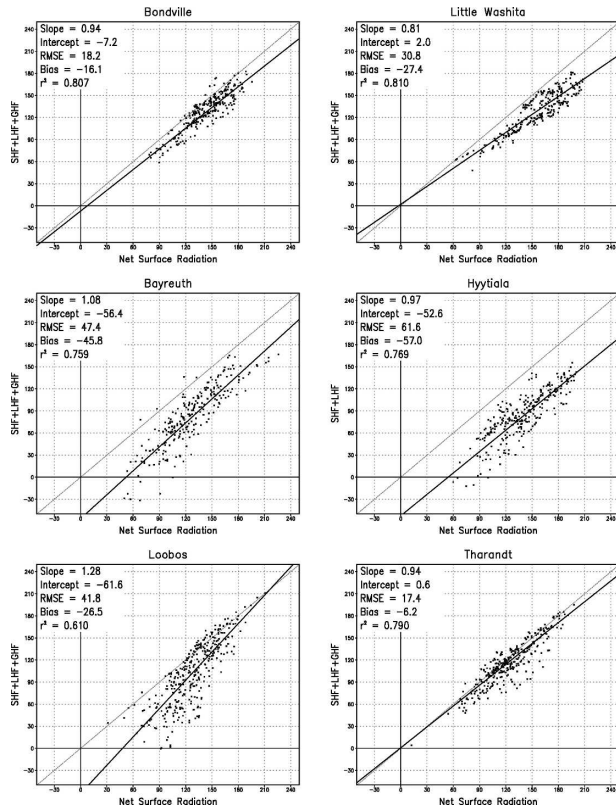


FIG. 3. As in Fig. 2 but for selected FLUXNET sites. Note Hyttiala lacks ground heat flux measurements.

tation datasets. The first is the multidecadal (1948–97), $1/4^\circ$ daily precipitation reanalysis of Higgins et al. (2000), which is based on rain gauge reports from over 10 000 sources in the United States. The second dataset consists of global monthly precipitation fields generated by Global Precipitation Climatology Project (GPCP; version 2; see Huffman et al. 1997). This global dataset, which combines in situ (gauge) and satellite measurements, covers the period 1979–2001 at a spatial resolution of $2.5^\circ \times 2.5^\circ$.

3. Observable analogs to “coupling strength”

In this study we use only data from cases W and S, as described in the introduction. The definition of the change in intraensemble coherence of model evapotranspiration (ET) from the control case to the case where subsurface soil wetness (SWet; wilting point = 0, saturation = 1) is specified, referred to as $\Omega_E(S) - \Omega_E(W)$ in G06 but here simply called $\Delta\Omega_E$, carries clear implications. It suggests that increased coherence must be the result of a strong functional dependence of ET on SWet. If there is no relationship between ET and

SWet, the specification of a particular time series of SWet as a boundary condition common to all ensemble members should have no statistically detectable effect on the ET time series. Of course, the land surface parameterizations in every one of the 12 GLACE models specify SWet as a term on the rhs of their respective equations for evapotranspiration [or more likely, for latent heat flux (LHF)]. LHF is not a univariate function, but also depends on other state variables (and parameters that are functions of state variables). The degree to which SWet specifically and uniquely determines LHF likely varies among models, geographically within a model, and even temporally at any grid box within a model, depending on the impacts of the other predictors in the equation.

In fact, one would expect to determine most strongly not the absolute LHF, but rather the partitioning of available energy between LHF and sensible heat flux (SHF). We may examine this effect through the normalized latent heat (NLH) defined as the ratio of LHF to net surface radiation. We can compare the functional dependence of NLH on SWet among models and to observations, as well as calculate values of Ω_{NLH} and $\Delta\Omega_{NLH}$.

Figure 4 shows scatter diagrams of NLH as a function of SWet for all members of the control ensemble from nine of the GLACE GCMs at the grid cell containing the center of the ARM region (some of the models did not provide the complete set of output or had other problems that precluded computing all of the necessary quantities for this part of the study). Six-day means are shown beginning at day 8 of the 92-day integrations (like those used in the calculations by K06 and G06). However, we display the 6-day means beginning every day through day 87—a total of 80 points per ensemble member instead of 14. This helps to show the evolution of the two terms in time for some of the models, but significance tests use the degrees of freedom for the smaller set of consecutive 6-day means. A few models show a very smooth and tight relationship between NLH and SWet (e.g., NSIPP) while others appear to have little relationship at all (e.g., CSIRO-CC3). Other contrasts exist. Some models span the entire range of SWet (e.g., GFDL) while others have a very limited range (e.g., CAM3) or a very uneven distribution (e.g., GFS/OSU). There is also a great deal of discrepancy in the ranges of SWet among models, although they all span most of the range of NLH.

The blue lines in the panels of Fig. 4 are a best fit to the scatter of points, based on 20 bins of equal population of points along the SWet axis spanning the range of SWet for that model and location. For each bin, the

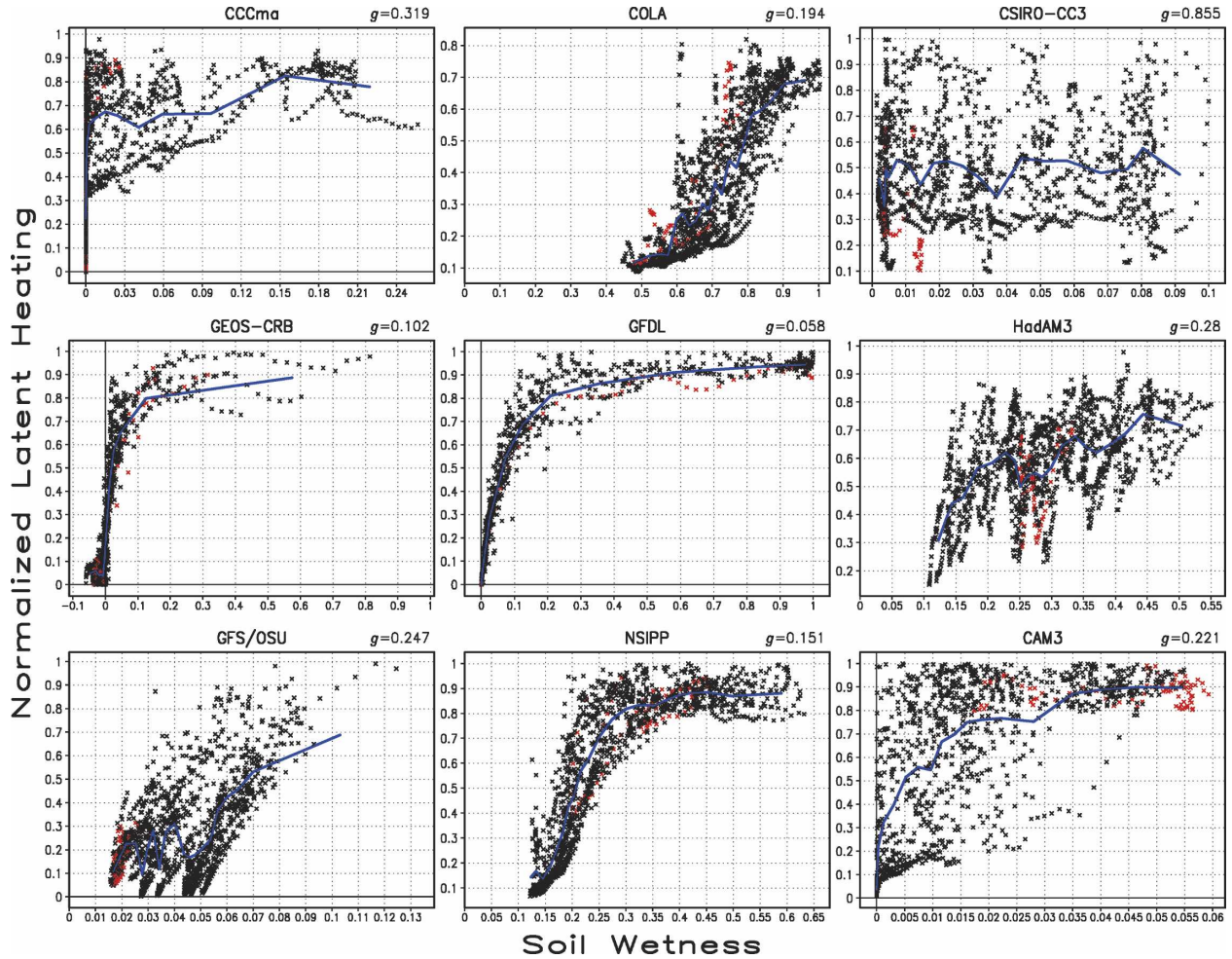


FIG. 4. Relationship of NLH to SWet in the 16 ensemble members of nine GCMs at the grid box encompassing the ARM Central Facility. Solid blue line is fit through the means of 20 bins of equal number of points. Red points show the ensemble member used as basis for fixed SWet integrations. Here g is a goodness-of-fit metric.

average value of NLH is calculated. The line connects those values. The limited sample size contributes to the zigzag nature of this line for some models. The advantage of this approach is that no a priori assumption is made regarding the functional relationship between NLH and SWet.

To match $\Delta\Omega_{NLH}$ to the degree of dependence of NLH on SWet, we need to quantify the strength of the functional relationship between the two quantities with a single value at each grid box. It is fairly easy to discern by eye from Fig. 4 which models exhibit a strong dependence of NLH on SWet and which do not, but we need an objective, quantitative means to do so. We estimate the strength of the functional relationship as a ratio. The numerator is the standard deviation of the LHF values in each bin i about the bin average, totaled over all bins:

$$s = \left(\frac{\sum_i \sum_{n_i} (NLH_{n_i} - \overline{NLH}_i)^2}{\sum_i n_i} \right)^{1/2}. \quad (1)$$

The denominator is the total range of the 20 bin-averaged NLH values:

$$R = \max(\overline{NLH}_i) - \min(\overline{NLH}_i). \quad (2)$$

The result is an estimate of “goodness of fit”:

$$g = \frac{s}{R}, \quad (3)$$

where g is a positive number whose value decreases as the fit improves. We have conducted Monte Carlo simulations showing that for data distributed in a Gaussian-random fashion on both x and y axes, g also has a

Gaussian distribution, and values of g below 0.36 are significant at the 99% level. The values of g for each model at the grid box encompassing the core of the ARM network are also shown in Fig. 4. Only CSIRO-CC3 fails to achieve statistical significance at this level.

The ensemble member chosen as the source for the fixed SWet runs is shown in the scatterplots of Fig. 4 by red symbols. This illuminates another shortcoming of the design of the GLACE experiment. The resulting value of $\Delta\Omega_E$, $\Delta\Omega_{NLH}$, and potentially $\Delta\Omega_P$ at any location for a given model may be a result of the random choice made in selecting the basis for specified SWet, especially for the majority of models that do not appear to span the entire range of possible SWet values during one seasonal integration. We can see from Fig. 4 that for this location the chosen SWet time series in the CAM3 model happened to be the wettest of all ensemble members. This may have depressed the estimate of $\Delta\Omega_{NLH}$ at this location. On the other hand, CCCma chose an anomalously dry case where the slope of the fitted curve is large and sensitivity is unusually high.

Ideally, to avoid this problem, the sensitivity experiments in GLACE would have been carried out 16 times, once with each control ensemble member as the source of specified SWet. This was an impractical demand to make on the modeling groups. Of course, through the averaging of $\Delta\Omega_E$ across the globe, this effect is averaged out, allowing the GLACE design to provide an accurate overall assessment of an individual model's coupling strength (see Fig. 6 in K06). In addition, at a given location, averaging across the models should filter out much of this source of error.

We also see from Fig. 4 that some models appear not to have a strict dependence of NLH on SWet, but a strong codependence on other factors. The CCCma, CSIRO-CC3, HadAM3, and GFS/OSU models (and COLA and CAM3 to a lesser extent) show evidence of this as oscillations or deviations in the track of points from the "best fit" envelope. We checked whether these models had a high incidence of drizzle that might drive a large proportion of evaporation to come from interception loss (which would occur independently of SWet), but that was not the case. Other factors must exert significant control on NLH (and LHF, not shown) in these models.

Figure 5 shows global maps of $\Delta\Omega_{NLH}$ for each model. There is a fairly strong agreement between $\Delta\Omega_{NLH}$ and $\Delta\Omega_E$ for each model (not shown), but generally $\Delta\Omega_E$ is larger. Shown in each panel is the global mean (land only, north of 60°S) value of $\Delta\Omega_{NLH}$. Consistent with the findings of G06, the GFDL model has the strongest $\Delta\Omega_{NLH}$ and the GFS/OSU model is the

weakest. Figure 6 shows the global distribution of the goodness-of-fit parameter g for each model. The shading is chosen so that statistically significant functional relationships of NLH on SWet are shown in shades of blue where they exceed the 99% confidence level, orange for confidence between 90% and 99%, and yellow for values below 90%. Shown in each panel are the fraction of the nonblank land area where confidence exceeds 99%, as well as the global mean of g and its spatial correlation with $\Delta\Omega_{NLH}$. Field significance is high for all models, and every model except CCCma has a statistically significant spatial correlation between g and $\Delta\Omega_{NLH}$, implying that the goodness-of-fit diagnostic is indeed relevant to coupling strength. Blank areas over land are either very dry (low soil wetness or low variance of soil wetness) or have no valid value of $\Delta\Omega_{NLH}$.

There exists a resemblance between the spatial patterns of the multimodel values of $\Delta\Omega_{NLH}$ and g , shown in Fig. 7. The global mean of g is 0.467 for the nine-model mean, and the spatial correlation between the two fields is -0.57 . It seems clear that once the noise from the original limited set of GLACE integrations is filtered out by aggregation, a firmer relationship is established between the lower branch of the land-atmosphere feedback loop and locally observable quantities. However, when we consider calculations of g and $\Delta\Omega$ based on LHF instead of NLH, the multimodel average shows a much higher spatial correlation between the global fields of -0.73 , explaining over half of the variance. The stronger connection for LHF than for NLH in the models is counter to what is suggested in the observations, as we will show later.

This exercise suggests that within the realm of weather and climate models we may relate the unmeasurable coupling indices $\Delta\Omega_{NLH}$ and $\Delta\Omega_{LHF}$ to an index that is not dependent on ensembling. This opens the possibility that we can quantify aspects of coupling strength between land and atmosphere in the real world, given a sufficiently large and high quality set of measurements over several seasons at locations of interest. Additionally, this relationship gives us a means to validate the coupling characteristics of these models, given the caveats mentioned earlier. At the very least, we can test whether these models simulate the correct ranges and sensitivities of surface fluxes and state variables.

Table 1 shows how the individual models compare to the observations of SWet, LHF, and g calculated for both NLH and LHF at Bondville, Little Washita, and the average of the ARM sites. Averaging over the ARM domain helps scale these observations to GCM grid scales, and avoid errors from local surface variabil-

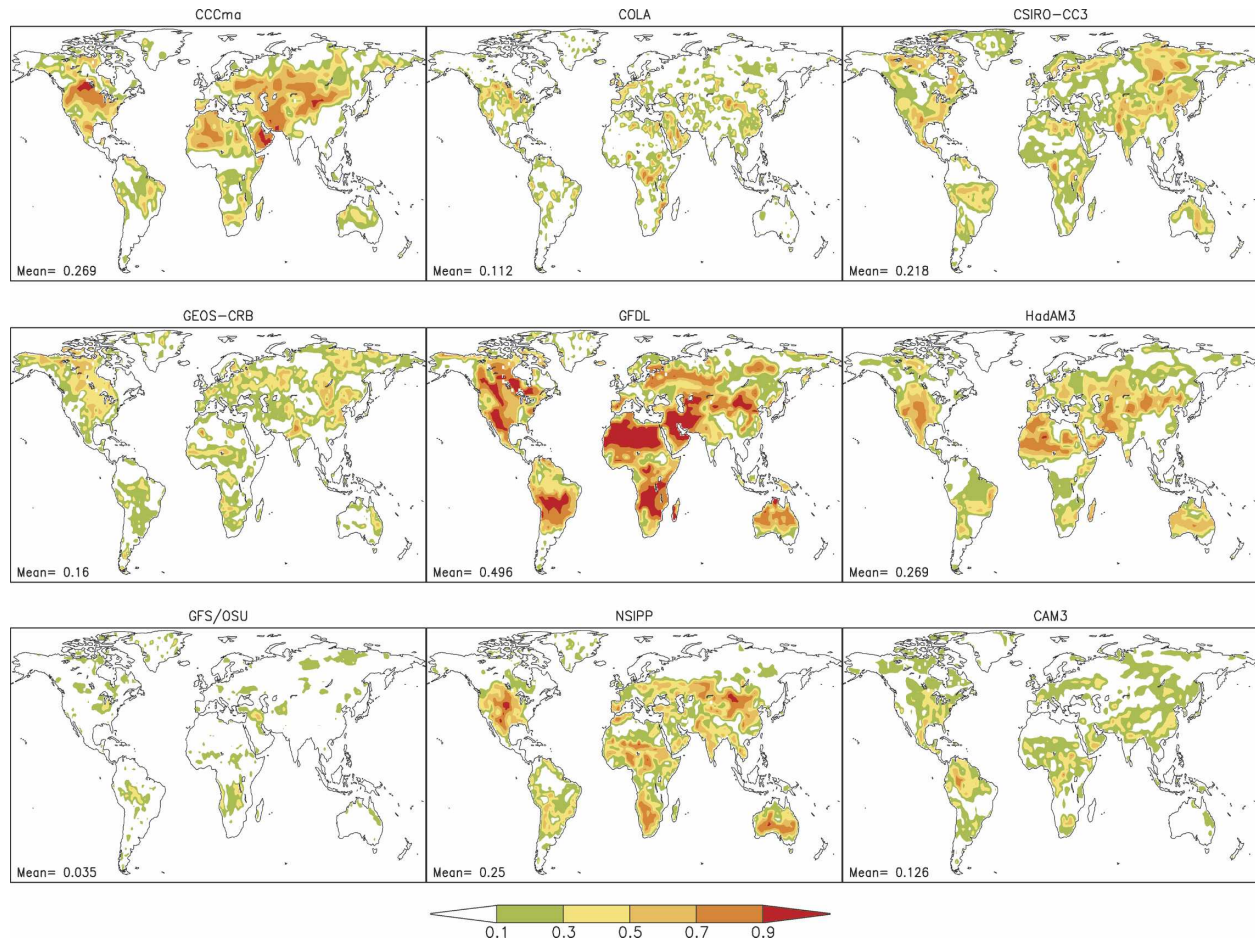


FIG. 5. The $\Delta\Omega_{NLH}$ for boreal summer in each model. Global mean (land only) value is shown in the bottom left corner of each panel.

ity and nonlinearities (Crow and Wood 2002). Although the HadAM3 and CCCma are consistently among the best models in terms of error for all the quantities shown, none of the models is especially impressive. The type of variability among models shown in Fig. 4 is typical for all of the locations examined, and none of the models is comfortably accurate in its representation of the observed relationships between SWet and NLH or LHF. As can be seen in Table 1 and Fig. 4, the models struggle to represent the correct distribution of soil wetness, and rarely come within 20% of the observed mean values of any quantity. It is also worth noting that in most cases the models have a better fit for the functional relationship of LHF on SWet than of NLH on SWet. The station data suggest the opposite is true. It seems that most of the models favor a stronger dependence of ET on SWet than for the partitioning of net radiation on SWet. The GFDL model bucks the trend in this regard but shows much too strong a relationship between SWet and surface

fluxes at all locations. CAM3 has the correct stratification of g at two of three locations, and has a much better goodness of fit than GFDL and most other models. The multimodel average (last column) ranks in the top four in 12 of 15 rows, giving further credence to its use as the best model-based representation of the real world.

In Fig. 8 we show how the observations behave over the ARM domain in terms of the relationship between NLH and SWet. Comparison to Fig. 4 shows just how different the models are from the observations. Note that only four years of measurements have gone into Fig. 8—no more than one-quarter the amount of data in the model plots. Thus Fig. 8 may underrepresent the observed range of SWet due to the small sample size and the possible lack of measurements at very low SWet. Nevertheless, the “best” models are NSIPP, which has the right goodness of fit but appears to put too much energy toward ET, and HadAM3, which overlaps the range of SWet and NLH rather well but

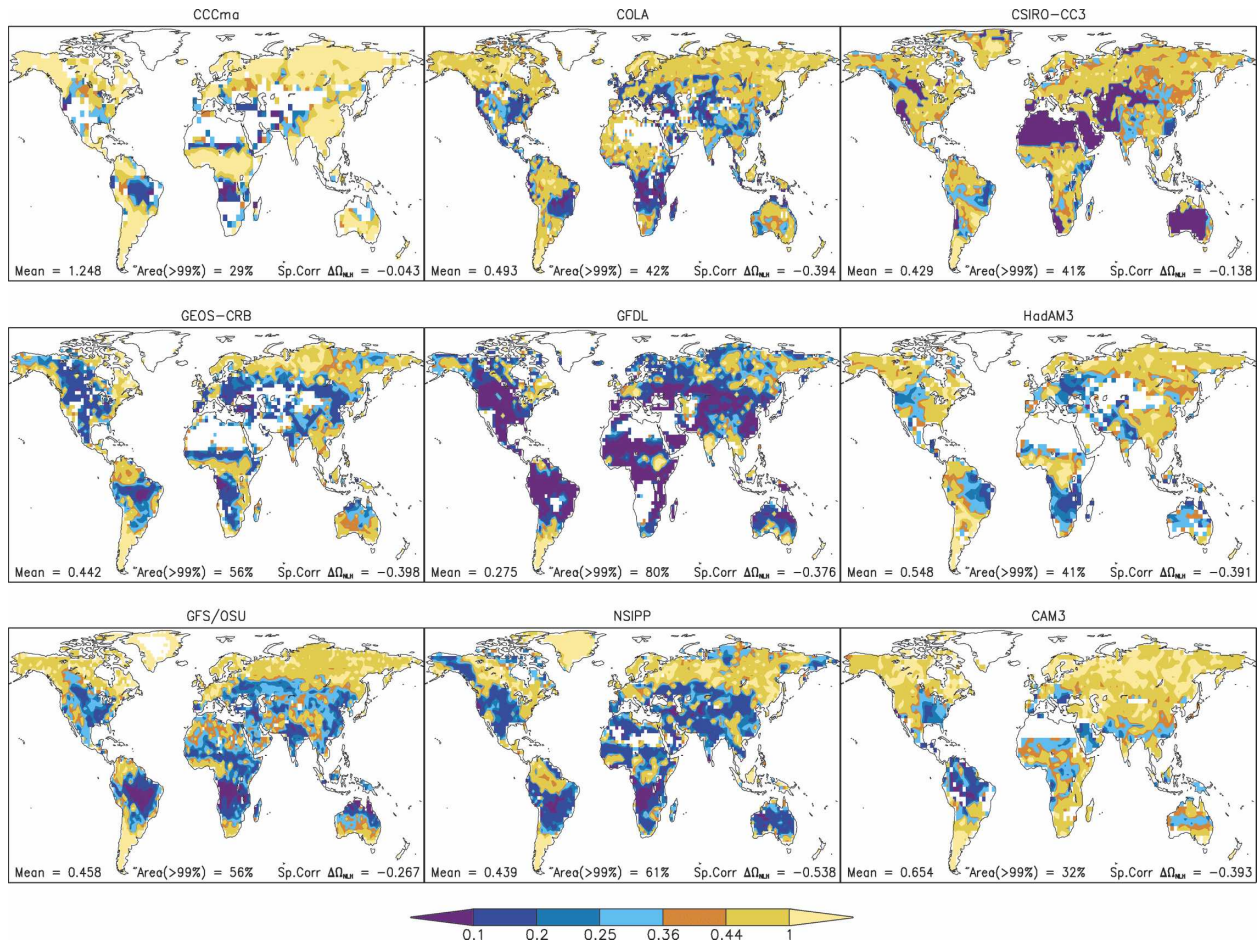


FIG. 6. As in Fig. 4 but for g . Also shown at the bottom center of each panel is the global spatial correlation between $\Delta\Omega_{NLH}$ and g for each model.

exhibits a slow mode of variability (the trains of points arcing up and down on a slight tilt) that increases spread and is not evident in the observations. The ARM data do show a positive correlation between NLH and SWet. The observed fit is much weaker in the observations than in the models. Because of the lack of observational data in very dry conditions, we cannot say whether that tail of the relationship follows an exponential curve like the GFDL or GEOS-CRB models, or an S-shaped curve like NSIPP and COLA. Overall, comparisons at the individual ARM sites, Bondville and Little Washita, portray a similar picture.

4. Other relationships with surface variables

Our comparison with observational sites is greatly restricted by the need for long time series of soil wetness measurements. However, there may be relationships between other more commonly measured surface

variables that we can use for validation and comparison. B04 found a number of striking associations among surface and lower-atmospheric quantities in the ECMWF reanalyses that can be used as a guide for this investigation. For instance, the relationship between surface SHF and SWet was found by B04 to be stronger than between LHF and SWet across several domains from the deep Tropics to boreal forests. This characteristic is largely borne out in the observations (Table 2). Only Elk Falls in the ARM network and Bondville show a significantly lower value of g for the relationship between SWet and LHF than between SWet and SHF. However, when we normalize by net surface radiation, the relationship reverses. Only at Elk Falls and Little Washita is the value of g appreciably lower for normalized sensible heat (NSH) than for NLH. At the same time, the goodness of fit increases going from LHF to NLH at every station, while it decreases going from SHF to NSH at all but three stations.

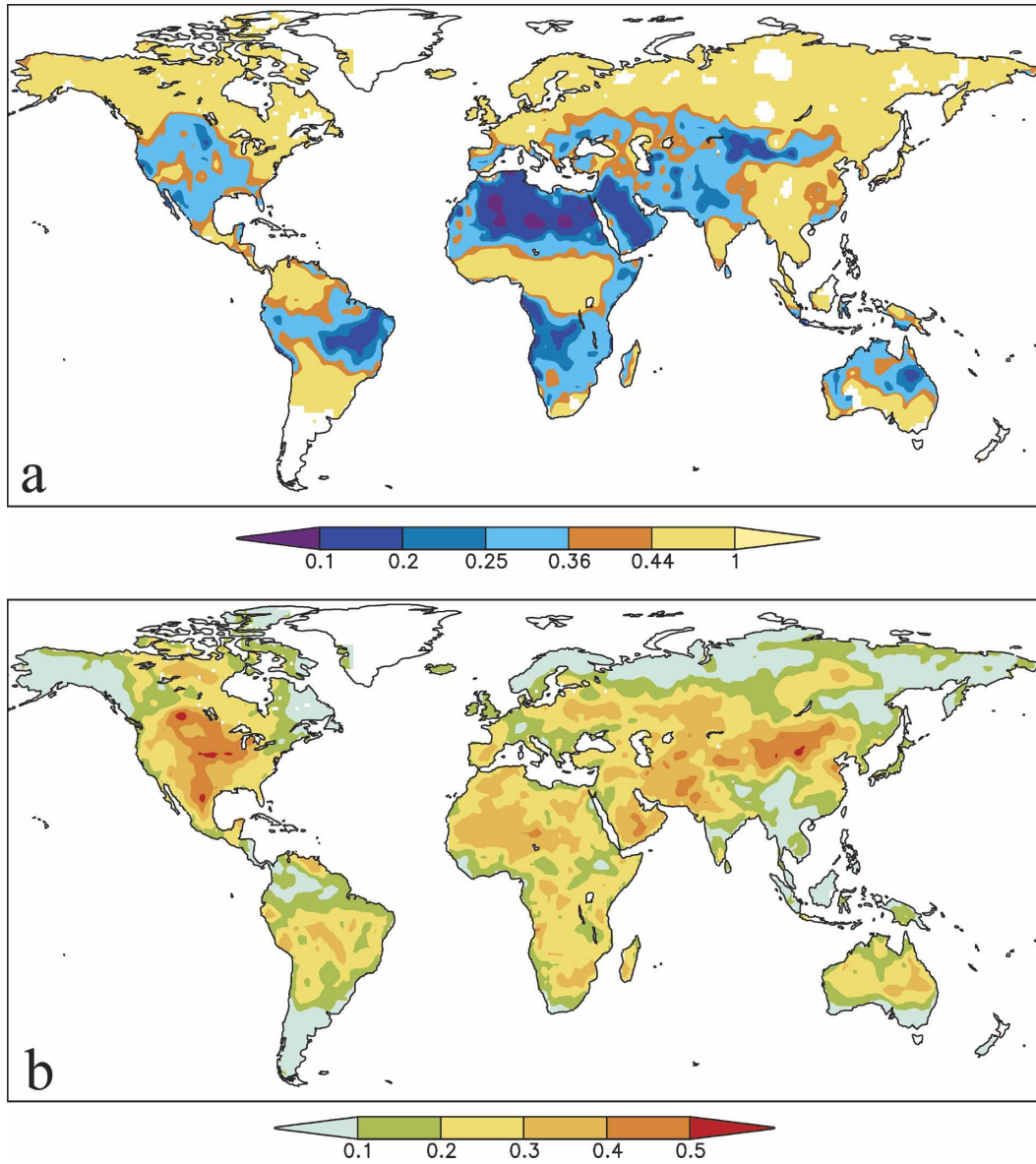


FIG. 7. The multimodel mean of (a) g and (b) $\Delta\Omega_{NLH}$.

The GLACE models have a very different behavior. Table 3 shows that there is a stronger functional relationship between LHF and SW_{et} than for SHF and SW_{et} for every model over the ARM region, with the exception of CSIRO-CC3, which has a very weak relationship to either. The same is nearly always true at individual sites. Values of g for the average of the ARM sites are also shown for comparison to the model grid box values. Every model except GFS/OSU shows a stronger relationship between NLH and SW_{et} than for NSH and SW_{et}, consistent with observations, but five of nine models show a degradation in goodness of fit going from LHF to NLH, and every model except

CCCma and CSIRO-CC3 has a tighter relationship between NSH and SW_{et} than between SHF and SW_{et}, contrary to the observed data. The model values of g for NLH are the same as in Fig. 4. The implication is that the GLACE models all have a fundamentally different (and perhaps wrong) interplay between soil wetness and surface fluxes, at least in this region.

One possible explanation for this behavior is that the GCMs emphasize a different factor controlling surface heat flux than does the real world. For example the Penman–Monteith equation and similar relationships that are widely used to parameterize evapotranspiration in land surface parameterizations has two main

TABLE 1. Comparison of observations, models, and multimodel average estimates of SWet (dimensionless), LHF ($W m^{-2}$), and goodness of fit of NLH and LHF to SWet for two North American FLUXNET locations and the average over ARM Extended Facility sites.

	Obs	CCCma	COLA	CSIRO-CC3	GEOS-CRB	GFDL	HadAM3	CAM3	GFS/OSU	NSIPP	M.M.
Bondville	g(NLH, SWet)	0.285	0.282	0.150	0.518	0.089	0.058	0.501	0.200	0.241	0.332
	g(LHF, SWet)	0.354	0.242	0.128	0.467	0.118	0.138	0.408	0.198	0.258	0.264
	SWet mean	0.757	0.218	0.515	0.149	0.388	0.437	0.392	0.006	0.144	0.269
	SWet range	0.573	0.585	0.907	0.141	1.124	0.990	0.399	0.047	0.308	0.314
	LHF mean	100.8	118.5	136.9	121.5	120.9	116.8	121.9	63.2	88.7	149.3
Little Washita	g(NLH, SWet)	0.214	0.319	0.207	0.958	0.125	0.055	0.349	0.239	0.219	0.138
	g(LHF, SWet)	0.294	0.253	0.183	1.173	0.114	0.094	0.296	0.265	0.203	0.120
	SWet mean	0.347	0.030	0.595	0.014	0.012	0.113	0.223	0.017	0.032	0.229
	SWet range	0.289	0.256	0.741	0.072	0.760	0.989	0.467	0.068	0.080	0.476
	LHF mean	57.0	68.5	72.1	86.5	26.5	42.6	85.4	88.0	39.6	77.8
ARM average	g(NLH, SWet)	0.201	0.319	0.194	0.855	0.102	0.058	0.280	0.221	0.247	0.151
	g(LHF, SWet)	0.427	0.253	0.186	0.988	0.085	0.103	0.284	0.291	0.223	0.146
	SWet mean	0.367	0.030	0.705	0.033	0.022	0.205	0.289	0.018	0.043	0.291
	SWet range	0.338	0.256	0.558	0.099	1.021	0.999	0.442	0.059	0.118	0.531
	LHF mean	104.2	68.5	79.5	74.7	35.8	67.6	95.4	107.8	42.4	108.5

terms: one based on potential evapotranspiration (effectively net radiation) and the other on the humidity gradient between the land surface and near-surface air. There is a lack complete information (namely, aerodynamic resistance) that would allow us to directly compare the relative magnitudes of each term for each model and for observations. However, the main com-

ponents of each term among the models and observations can be compared.

Figure 9 shows the frequency distribution of net radiation (top), the difference between actual and saturation specific humidity (middle), and the temperature (bottom) for the ARM region. Observed distributions are shown by bars. The vertical lines show the range among the GCMs in each bin, with the marker indicating the nine-model mean frequency. Note that the GCMs have a reasonable distribution of net radiation (only one GCM has a distinct high bias, which principally affects the ranges in the panel). However, all GCMs have a propensity for unrealistically large spe-

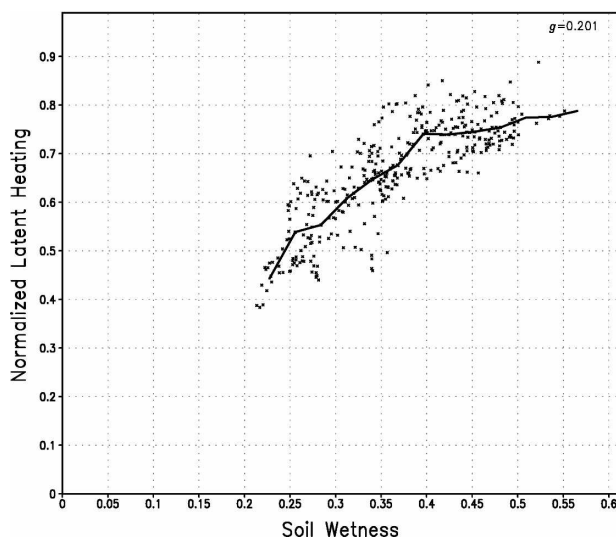


FIG. 8. As in Fig. 3 but for observed average over ARM Extended Facility sites.

TABLE 2. Observed goodness of fit between various surface flux variables and SWet at individual sites in North America.

	g(LHF, SWet)	g(SHF, SWet)	g(NLH, SWet)	g(NSH, SWet)
E8—Coldwater	0.313	0.283	0.247	0.448
E22—Cordell	0.350	0.282	0.284	0.310
E7—Elk Falls	0.330	0.368	0.316	0.298
E19—El Reno	0.763	0.523	0.640	0.652
E2—Hillsboro	0.508	0.514	0.353	0.848
E13—Lamont	0.496	0.472	0.438	0.476
E20—Meeker	0.330	0.199	0.197	0.193
E18—Morris	1.057	0.631	0.563	0.844
E12—Pawhuska	0.339	0.173	0.215	0.451
Bondville	0.354	0.419	0.285	0.705
Little Washita	0.294	0.199	0.214	0.170

TABLE 3. As in Table 2, but for models and observations for the average over ARM Extended Facility sites.

	$g(\text{LHF}, \text{SWet})$	$g(\text{SHF}, \text{SWet})$	$g(\text{NLH}, \text{SWet})$	$g(\text{NSH}, \text{SWet})$
Observations	0.427	0.232	0.201	0.248
CCCma	0.253	0.337	0.319	0.342
COLA	0.186	0.260	0.194	0.204
CSIRO-CC3	0.988	0.743	0.855	1.045
GEOS-CRB	0.085	0.176	0.102	0.118
GFDL	0.103	0.151	0.058	0.126
HadAM3	0.284	0.371	0.280	0.308
CAM3	0.291	0.335	0.221	0.228
GFS/OSU	0.223	0.295	0.247	0.245
NSIPP	0.146	0.219	0.151	0.153

cific humidity depressions (and thus low relative humidity). In fact, nearly a third of the days in the typical model have values above the observed range. This appears to be a result of excessively high temperatures in the GCMs (bottom panel). Whereas there are no occurrences of surface air temperatures warmer than 33°C in the observational data (a mean over the ARM stations), individual models simulate anywhere from 10% to 55% of their days with mean temperatures above this level. Given the highly nonlinear increase of saturation specific humidity with temperature at these high values, it appears likely that the GCMs' evapotranspiration is too strongly driven by humidity gradients (e.g., by the vapor pressure deficit term in the Penman equation), and thus responds relatively weakly to variations in net radiation.

This result, though striking, is for only one location. Do the models show this apparent overdependence on vapor pressure deficit globally? Figure 10 shows the ratio of g calculated for LHF versus g for NLH. Blue areas show where the models overall have a stronger dependence on SWet by LHF than NLH. This includes most of the midlatitudes, including the areas in Europe and North America where this study has observational data. Table 4 compares the global mean values of g and the area where LHF has a stronger dependence on SWet than does NLH for the GLACE models. No model shows a dominance for NLH in the global mean, and only the GFDL and CSIRO-CC3 models show $g(\text{NLH}, \text{SWet})$ dominating over a majority of the land area. Similar comparisons between $g(\text{LHF}, \text{SWet})$ and $g(\text{SHF}, \text{SWet})$ show that most models have a stronger dependence of LHF on SWet than SHF on SWet.

So overall, most models do not show the relative dependencies of surface fluxes on soil wetness that are suggested by B04 or the limited observational data available. This may result, at least in part, from biases in simulated vapor pressure deficit. These flaws, however,

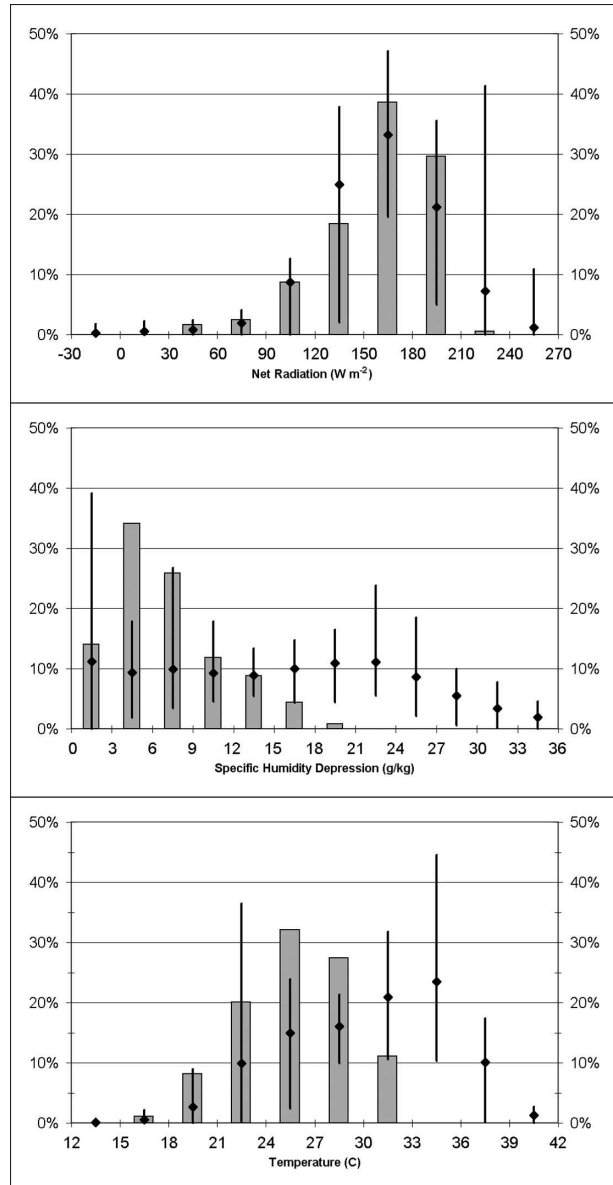


FIG. 9. (top) Categorical frequency of occurrence of net radiation, and (middle) the difference between actual and saturation specific humidity and (bottom) temperature over the ARM region for observations (bars), and the mean of the GCMs (markers). Vertical lines span the range of models for each bin.

do not necessarily invalidate the pattern and degree of land-atmosphere coupling found by K04. B04 contends that the strong relationship between SHF and SWet is not necessarily direct, but through the strong interactions each have with the height above ground (in pressure coordinates) of the lifting condensation level (P_{LCL}). The proposed mechanism is that SWet exerts a strong control on P_{LCL} through its effect on the near-surface dewpoint depression, which then determines

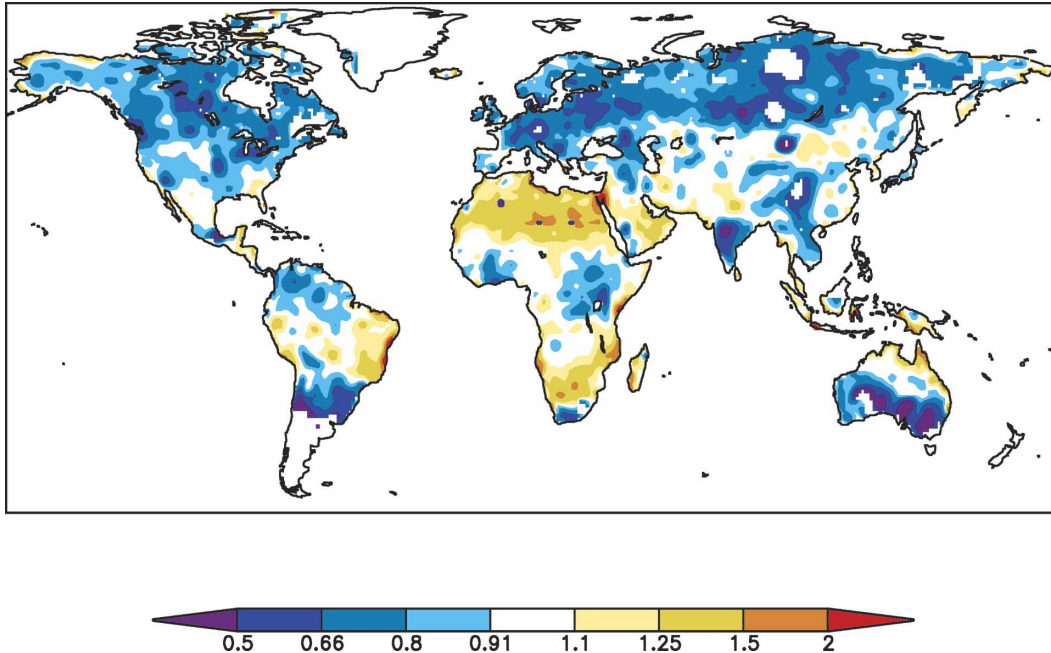


FIG. 10. Ratio of the multimodel mean of $g(\text{LHF}, \text{SWet})$ to $g(\text{NLH}, \text{SWet})$.

the size of the available heat reservoir in the mixed layer, and thus the rate of SHF that can be sustained. A nearly uniform heating rate of the boundary layer of 3.8 K day^{-1} was found by B04 for SHF in the ECMWF model forecasts, yielding a linear relationship between the mass of air in the boundary layer and the sensible heat flux rate when averaged over 5-day intervals. So SWet may impact cloud processes, and thus precipitation, via both LHF and SHF.

We calculate the observed and model relationships between P_{LCL} and SWet following the approximation for P_{LCL} based on near-surface temperature and hu-

midity from Bolton (1980). The average of the observations over the ARM region (Fig. 11) shows a fairly strong relationship similar to B04. The range of soil wetness is limited in this area, so the tails of the distribution for very wet and dry soil conditions cannot be seen. The models (Fig. 12) exhibit an assortment of behaviors, but all GCMs except GFDL have a high bias in P_{LCL} and most have a clear negative correlation between SWet and P_{LCL} . The variety is striking. GFDL, for example, has a very tight connection between P_{LCL} and SWet, while some other models show a rather weak relationship between these variables (e.g., HadAM3 or GFS/OSU) or no relationship at all (e.g., CCCma or CSIRO-CC3). The models stratify just as they did for the other goodness-of-fit relationships. It appears that many of these GCMs do not simulate the proper coupling between surface moisture and the cloud base. The positive biases in cloud-base height are consistent with biases toward low relative humidity shown in Fig. 9, suggesting a connection between these errors in the models.

What about the relationship between P_{LCL} and SHF? We can now introduce the data from the EUROFLUX sites into the validation exercise. The models in GLACE did not report SHF, but we can deduce the term $\text{SHF} + \text{GHF}$ (ground heat flux) from LHF and net radiation. In Table 5 the observations are shown for the implied heating rates and the r^2 with P_{LCL} using both SHF and $\text{SHF} + \text{GHF}$ where available to provide

TABLE 4. Global mean values from models and the multimodel mean of goodness of fit, the ratio of the global means of goodness of fit, and the fraction of global land surface grid points where the dependence of LHF on SWet is stronger than for NLH on SWet.

	$g(\text{LHF}, \text{SWet})$	$g(\text{NLH}, \text{SWet})$	Ratio	Area $g(\text{LHF}, \text{SWet}) < g(\text{NLH}, \text{SWet})$
CCCma	0.483	1.248	0.39	85%
COLA	0.443	0.493	0.88	62%
CSIRO-CC3	0.413	0.429	0.96	42%
GEOS-CRB	0.357	0.442	0.81	67%
GFDL	0.229	0.275	0.83	27%
HadAM3	0.496	0.548	0.91	55%
CAM3	0.556	0.654	0.85	59%
GFS/OSU	0.377	0.458	0.82	70%
NSIPP	0.383	0.439	0.87	54%
MM	0.400	0.447	0.89	69%

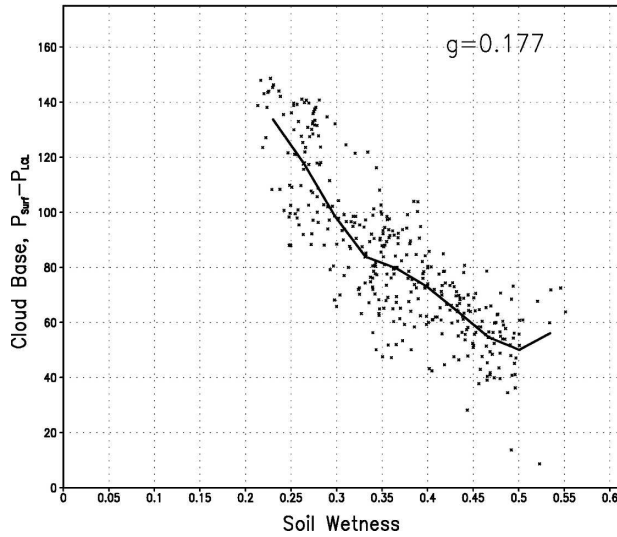


FIG. 11. As in Fig. 7 but for the relationship between height of cloud base (hPa) and SWet.

a means of translation to the results of B04. Over the ARM sites the heating rate deduced in B04 appears quite appropriate, but for the other FLUXNET sites a range of heating rates from 2.9 to 6.0 K day⁻¹ is apparent. Inclusion of GHF in the calculation tends to increase the slope, and thus the derived heating rate, and curiously also improves the fit of the linear regression in most cases.

Large differences in the value of r^2 between models and observations suggest that those models do not represent the relative importance of SHF as a source of boundary layer heating (or cooling) compared to other thermodynamic processes such as radiative cooling, thermal advection, diffusion, and dry and moist convective processes. However, a high value of r^2 does not guarantee a correct heating rate, because even if a particular model is producing a good simulation of SHF, the other heating terms in the boundary layer may be amiss. Table 5 suggests that while some models clearly

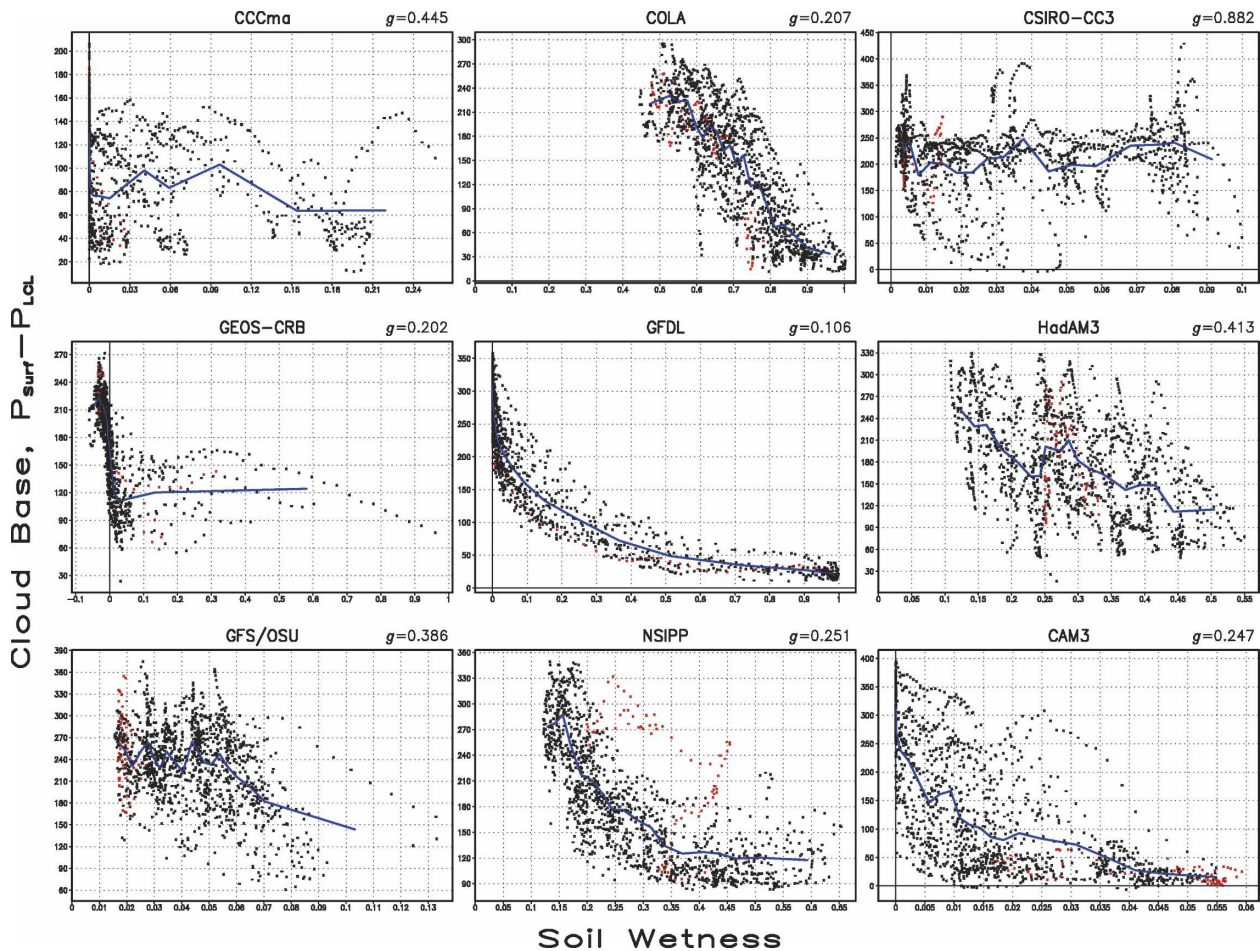


FIG. 12. As in Fig. 3 but for the relationship between height of cloud base (hPa) and SWet.

TABLE 5. Comparison of the percentage of explained variance between SHF and P_{LCL} , and the derived boundary layer heating rates ($K \text{ day}^{-1}$) for observations and models for the ARM region average as well as all available FLUXNET sites. The right column shows the results for the multimodel mean. The bottom rows show the average across all locations.

		Obs (SHF)	Obs (SHF+GHF)	BMRC	CCCma	COLA	CSIRO-CC3	GEOS-CRB	GFDL	HadAM3	CAM3	GFS/OSU	NSIPP	Model avg
ARM	r^2	0.59	0.63	0.03	0.70	0.72	0.21	0.37	0.74	0.65	0.78	0.36	0.49	0.51
	Heating rate	3.7	4.1	0.7	6.1	3.2	1.9	5.2	3.7	2.1	3.2	2.8	3.3	3.2
Bondville	r^2	0.27	0.22	0.03	0.58	0.89	0.60	0.55	0.73	0.70	0.69	0.57	0.00	0.53
	Heating rate	4.1	4.8	-0.4	3.9	4.9	2.2	3.6	3.2	3.1	3.4	3.4		3.0
Little Washita	r^2	0.59	0.65	0.03	0.70	0.58	0.16	0.37	0.72	0.69	0.76	0.37	0.53	0.49
	Heating rate	2.9	4.1	0.7	6.1	2.5	1.6	5.7	4.0	2.7	3.7	3.2	3.6	3.4
Bayreuth	r^2	0.40	0.51	0.01	0.31	0.09	0.52	0.00	0.00	0.60	0.61	0.18	0.15	0.25
	Heating rate	5.8	5.8	0.4	5.8	-2.0	2.2			4.0	3.9	5.3	3.5	2.9
Hyytiälä	r^2	0.55	N/A	0.46	0.30	0.61	0.81	0.58	0.60	0.76	0.58	0.25	0.36	0.53
	Heating rate	4.5	N/A	3.7	4.9	8.6	4.1	7.7	5.4	5.7	7.0	6.2	7.3	6.1
Loobos	r^2	0.51	0.57	0.27	0.38	0.63	0.39	0.46	0.32	0.65	0.62	0.21	0.17	0.41
	Heating rate	6.0	7.2	-5.3	7.6	4.0	3.5	6.8	10.7	3.2	5.3	7.8	-14.4	2.9
Tharandt	r^2	0.39	0.49	0.01	0.66	0.49	0.38	0.00	0.23	0.45	0.61	0.14	0.15	0.31
	Heating rate	3.1	3.7	0.4	7.5	3.6	2.2		5.5	2.9	3.4	4.3	3.5	3.7
Avg	r^2	0.47	0.52	0.12	0.52	0.57	0.44	0.33	0.48	0.64	0.66	0.30	0.26	0.43
	Heating rate	4.3	4.9	0.0	6.0	3.5	2.5	5.8	5.4	3.4	4.3	4.7	1.1	3.6

do better than others, none is without problems. Yet again, the multimodel average gives the most reasonable overall representation of variations in heating rates and correlations among the stations, although there are still biases in evidence.

Comparison of the observed relationships between surface and near-surface state variables, fluxes, and atmospheric parameters to those presented in B04 with forecasts from the ECMWF model, which did not participate in GLACE, shows that the ECMWF model has too little spread in many of the scatter diagrams. This resembles the GFDL model, which has the strongest coupling between SWet and surface fluxes in GLACE. The implication is that the ECMWF model might have

a similarly strong coupling between land and atmosphere. Another interesting aspect of the GFDL model is that over the ARM area it is the only model to show a clearly bimodal distribution of soil wetness (evident in Fig. 4). D'Odorico and Porporato (2004) argued that this can be a result of feedbacks between soil wetness and precipitation, which G06 showed to be strongest in the GFDL simulations.

5. Land-atmosphere coupling and precipitation memory

As discussed in the introduction, validation of the coupling strengths quantified in GLACE is difficult be-

cause direct large-scale observations of land–atmosphere feedback do not exist. We can, however, derive certain diagnostic quantities from large-scale observations that are tied indirectly to the feedback. These quantities, if examined with caution, allow an indirect evaluation of modeled coupling strength.

The two diagnostic quantities examined in this section are those described in detail by Koster et al. (2003) and Koster and Suarez (2004, hereafter referred to as K03 and KS04, respectively). Both K03 and KS04 followed the same analysis strategy: (a) a feature of interest—hypothesized as being related to land–atmosphere feedback—is identified in the observational data record; (b) the feature is then sought and identified in a full GCM simulation; (c) the GCM simulation is repeated with all land–atmosphere feedback artificially removed, and the absence of the feature is noted. The final two steps unequivocally identify land–atmosphere feedback as the source of the feature of interest *within the GCM*. Given the feature’s presence in the observations, we are left with two possible conclusions: either land–atmosphere feedback does indeed occur in nature, or the presence of the feature in both the observations and the model is coincidental.

The features identified by K03 and KS04 involve the spatial patterns of precipitation autocorrelation over the conterminous United States and the area-averaged conditional expected value of monthly precipitation following extreme precipitation months. Each feature is discussed here in the context of the GLACE results.

a. Patterns in the temporal correlations of precipitation

K03 speculated that land–atmosphere feedback, if it exists, should be reflected in the temporal correlations of precipitation. The idea is simple. If feedback operates in nature, an anomalously high precipitation event during one week should lead to high evaporation rates and thus high precipitation rates in subsequent weeks, strengthening the temporal correlation. K03 focused their analysis of the correlations on the continental United States, for which a precipitation dataset of acceptable length and quality is available (Higgins et al. 2000). Fifty years of daily July precipitation data, both from the observations and from the NSIPP-1 GCM (with or without enabled feedback), were aggregated to 5-day, or pentad, precipitation totals. Correlations were then computed between twice-removed pentads—that is, between the precipitation anomalies for 1–5 July and 10–15 July, between those for 6–10 July and 16–20 July, and so on. Correlations between consecutive pentads were not considered because these are overly influenced by storms that straddle the time divisions. A

statistically significant signal appeared in the observations for July and August. The NASA Seasonal-to-Interannual Prediction Project (NSIPP-1) GCM captured the overall shape of this signal, but significantly overestimated its magnitude. When feedback was disabled the correlations in the GCM essentially disappeared. Thus, feedback was responsible for the correlation signal in the GCM.

Using the pentad precipitation rates from the 16 Julys in experiment W (the free-running GLACE experiment, with no specification of surface states), we computed the correlations between twice-removed pentads for each GLACE model. (As in K03, the observational and model data fields were aggregated to the same $2^\circ \times 2.5^\circ$ resolution and treated with a 3-point filter prior to the calculation of the statistics.) The correlation fields are very noisy—not necessarily because the models are poor, but because the number of truly independent data pairs contributing to the correlation calculation for each model is small, of order 30. Still, several models show a rough indication of positive correlation in the center of the continent. For presentation here (Fig. 13), we filter out some of the sampling error by averaging the correlations across the continental United States and presenting the averaged results, for each of the three simulation months (June, July, and August), in histogram form. Individual models vary the exact location of the North American hotspot, so we use a large averaging area at the expense of reducing the values in the histograms.

In each panel of Fig. 13, the means for the observations are shown as dotted histogram bars. The observations show a maximum of correlation in July, a smaller amount in August, and a correlation in June that is close to zero. The models, as expected, show a range of behavior, with some models strongly overestimating the correlation (e.g., GFDL, CCCma) and others strongly underestimating it (e.g., GFS/OSU). In general, the models do not capture the observed seasonal cycle of the correlation.

Of course, given the pervasive sampling error, these results are hard to interpret, even with the spatial averaging. For reliable estimates of precipitation autocorrelation—particularly regarding nuances in seasonal and geographical distribution—hundreds of seasons should be examined, not just the 16 examined here. Still, the multimodel results shown at the bottom of the figure are encouraging. When sampling error and even model error is smoothed out further by averaging the spatially integrated values across the 12 models, the results for July and August are remarkably close to the observed results. The models still strongly overestimate

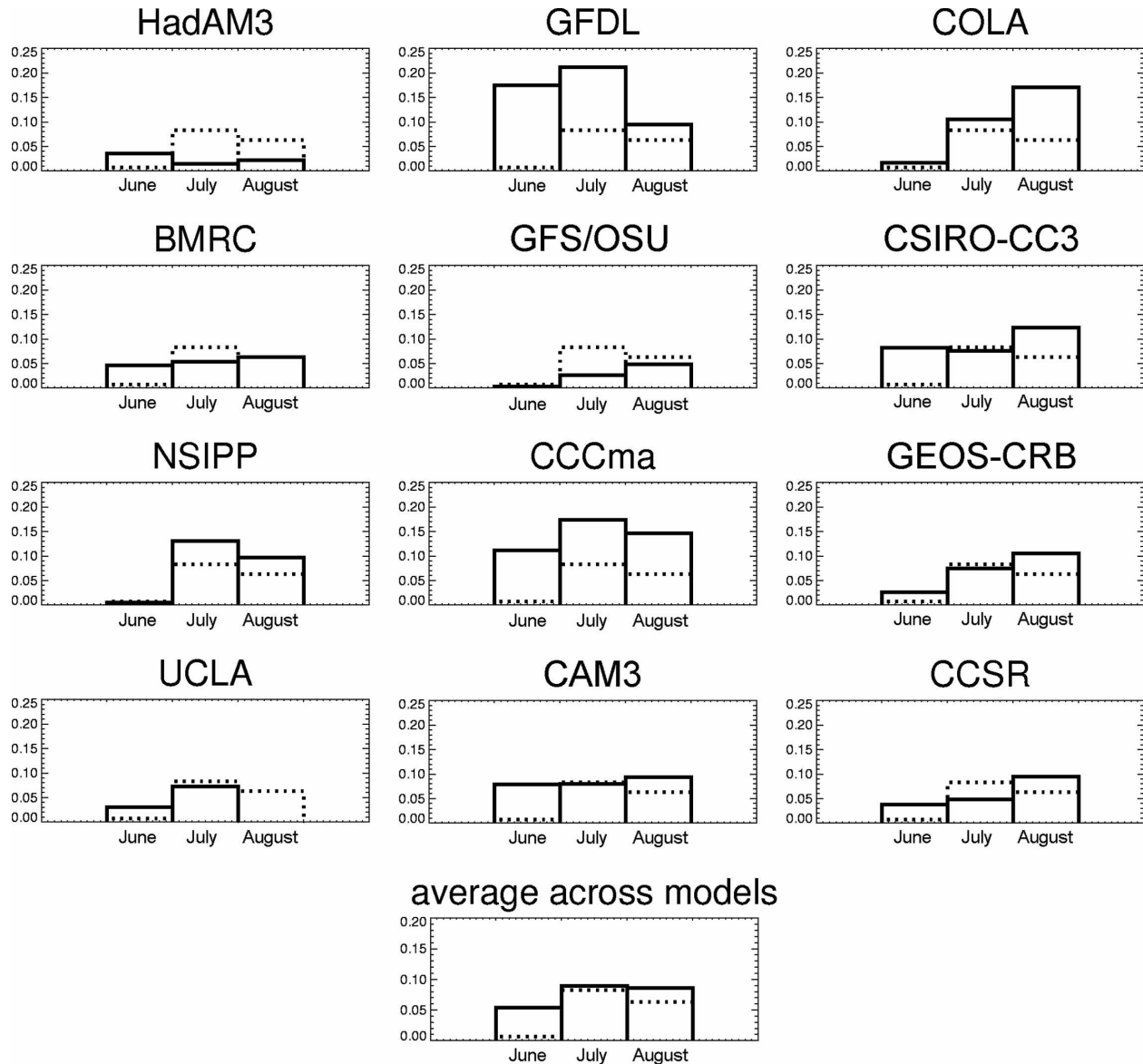


FIG. 13. Correlations between twice-removed 5-day precipitation totals averaged across the continental United States, as estimated from GLACE control ensemble output for each model (solid lines) and for observations (dashed lines).

correlations in June, though they correctly identify June as the weakest month for the correlations.

b. Conditional expected values of rainfall across midlatitude land

In KS04, observed monthly data were analyzed to determine the conditional expected value of a monthly precipitation anomaly given that the anomaly in a preceding month (one, two, or three months beforehand) was of a certain sign and magnitude. To increase the sample space and thereby allow meaningful distinctions between computed probability density functions

(PDFs), ergodicity was assumed: monthly precipitation totals in all grid cells covering midlatitude land (30°–60°N) were standardized and included in the construction of conditional probability distributions. To standardize the data, each monthly rainfall had the local mean subtracted from it, and the resulting anomaly was divided by the local standard deviation. The observed conditional expected values are statistically distinct. When the observed rainfall at a given location is in the lowest 20% (i.e., the lowest quintile) of all rainfalls at that location, the rainfall there in the following months also tends, on average, to be reduced. Similarly,

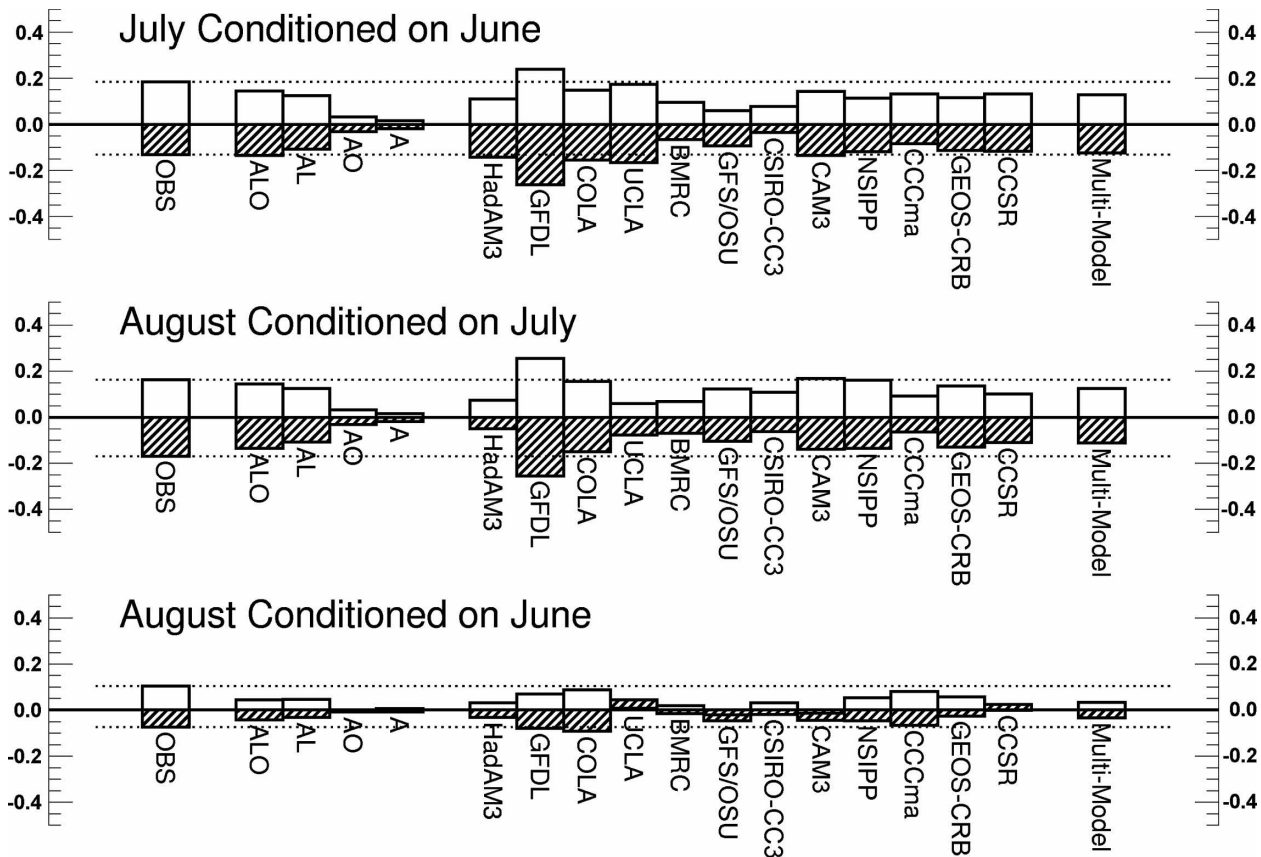


FIG. 14. Conditional expected mean of standardized precipitation anomaly given an antecedent monthly anomaly in the topmost quartile (clear bars) and in the bottommost quartile (striped bars). Results are shown for observations, the individual models, and the multimodel average. Results from KS04 are also shown: ALO refers to an AGCM run with atmospheric, land, and ocean variability acting; AL to a run with only atmospheric and land variability acting; AO to a run with only atmospheric and ocean variability acting; and A to a run with only atmospheric variability acting.

monthly rainfalls in the highest quintile tend to lead to higher-than-average rainfall rates in subsequent months.

KS04 then examined the precipitation rates generated in GCM simulations. The observed conditional expectations are reproduced by the GCM when land-atmosphere feedback is enabled, but they are destroyed when the feedback is artificially disabled. The effect of ocean variability on the signal is relatively small (see below). Thus, the GCM suggests that the observed conditional expectations are a signature of feedback.

The GLACE data allow the quantification of conditional expectations across a number of GCMs for comparison with the observations. Precipitation rates from the 16 Junes, Julys, and Augusts of the control ensemble (case W) were processed onto the same horizontal grid and then used to generate conditional PDFs following the strategy of KS04, with two slight modifications: (a) instead of binning the monthly rates into quintiles, which is difficult with 16 values, we binned

them into quartiles instead, and (b) rather than averaging the one-month-lagged results across the months studied, we separately examine July rainfall conditioned on June rainfall and August rainfall conditioned on July rainfall. Results are shown in Fig. 14. The observations (Huffmann et al. 1997) show that if June rainfall is in the top quartile, the standardized July rainfall will have an expected value of 0.2 (the unshaded, positive histogram bar). If, on the other hand, June rainfall is in the bottom quartile, the expected value of standardized July rainfall will be about -0.13 (the crosshatched negative bar). The results from the various models are generally similar in magnitude, but they still vary, with some models (notably GFDL) producing larger values and some others (BMRC, GFS/OSU, CSIRO-CC3) producing lower values. Results for August conditioned on July are similar. The averages of the conditional expectations across the models (the final bars in each panel) are close to, but slightly lower than, the observed values. For August conditioned on

June, the conditional expectations are greatly reduced, especially for the models. The multimodel averages of the conditional expectation for the two-month-lagged case are considerably less than the observed values.

The proper interpretation of the comparison between observations and models in Fig. 14 requires the careful consideration of ocean impacts. The conditional expectations for the observations may partially reflect an influence of sea surface temperature (SST) anomalies that span the summer season and that are different from year to year. Rainfall in the models cannot be similarly influenced, since all ensemble members utilize the same SST distribution. Again, the KS04 study suggests that land feedbacks dominate the signal. This is shown graphically in Fig. 14 by the histogram bars labeled ALO (for a control simulation, in which the atmosphere, land, and ocean all contribute to precipitation variability), AL (for a simulation in which the ocean's contribution is artificially suppressed), AO (for a simulation in which the land's contribution is artificially suppressed), and A (for a simulation in which the contributions of both the land and the ocean are suppressed). These are the original KS04 results: they are based on quintiles, and they are averaged across the five months that KS04 studied. KS04 concluded that land feedback dominates the diagnostic because without it (simulations AO and A), the conditional expectations are close to zero. Nevertheless, the histograms indicate that the ocean does have a nonnegligible impact. A comparison of the results for simulations ALO and AL (for the one-month-lagged cases) suggests that if the conditional expectations for the observations were not influenced by SSTs, the observational results might be reduced to about 90% of their plotted values. Considered in that light, the multimodel conditional expectations—at least for the one-month-lagged case—are seen to be very close to those from the observational data.

6. Discussion

We have revisited the output from the participating models of the GLACE experiment, which quantified the strength and distribution of land–atmosphere coupling within 12 GCMs and estimated a model-independent global distribution of land–atmosphere coupling. The results of K04, K06, and G06 are based on properties of the individual model ensembles. The present study attempts to validate to the fullest extent possible the behavior of the GLACE models with in situ observations. We look separately at the relationship between local surface properties and fluxes, and at the memory signal evident regionally in precipitation.

The g parameter (the goodness of fit of the empirically fitted dependence of NLH on SWet) correlates well with $\Delta\Omega_{\text{NLH}}$ and is thus considered an observable metric for a critical element of land–atmosphere coupling—the link between the soil moisture variations and surface fluxes. (The g parameter can be derived from observations wherever soil moisture measurements and flux towers are collocated, whereas $\Delta\Omega_{\text{NLH}}$ is a property of the ensembling of model integrations and is thus intrinsically unobservable.) Unfortunately, there are very few locations where contemporaneous measurements of surface fluxes and SWet have been collected over a sufficiently long period to provide statistically stable relationships. The ARM Extended Facilities and a subset of FLUXNET sites do provide sufficient data. At these locations, we find that individual models often poorly validate with regard to simulations of SWet, NLH, and the relationship between the two, but the multimodel average validates better. We also find that the models show a stronger relationship between LHF and SWet than between NLH and SWet, whereas observations show the reverse, suggesting at first glance that there may be some problems with the flux parameterizations in today's land surface schemes. Further investigation shows, however, that in the region studied (the ARM region), all of the models simulate excessively warm temperatures and unrealistically low daytime relative humidity, thereby reducing the relative impact of net radiation variations on the surface fluxes.

B04 provides a set of relationships, found within the ECMWF model, that allow us to extend the analysis to other variables such as SHF and P_{LCL} —variables that are measured or can be estimated at FLUXNET sites where SWet is not recorded. B04 and field data suggest that the relationship between SHF and SWet is usually stronger than that for LHF and SWet, but the GLACE models do not exhibit that characteristic. (The models do agree with observations that NLH and NSH have similar relationships with SWet, with NLH being slightly stronger.) Likewise, the relationship between SWet and P_{LCL} and between SHF and P_{LCL} found by B04 is generally borne out in the observations, but poorly represented by many of the GLACE models. Most models simulate too high a P_{LCL} , which is consistent with the excessive simulated surface temperatures. Most GCMs appear not to simulate properly the coupling between the land surface and atmospheric boundary layer in midlatitude summer. Several GLACE models show too weak a relationship, and the ECMWF model of B04 along with a few of the GLACE models appear to be too strongly coupled. Thus, perhaps, the multimodel estimate of land surface coupling strength

is not an unreasonable approximation of reality. It should be noted that the results of B04 were based on data from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40), whereas the GLACE models were not constrained by data assimilation. Nudging of the state variables would not affect the calculation of fluxes, but could limit the range of SW_{et} or alter the apparent relationship between SHF and P_{LCL} , since P_{LCL} is a function of near-surface temperature and dewpoint. It is unclear what affect this might have on the apparent coupling strength of the ECMWF model. Stated another way, the GLACE models may appear poor in this comparison not because the parameterizations underlying land-atmosphere coupling are poor, but because biases in the climate model shift the model climates into unrealistic regimes at the validation sites.

Large-scale relationships for precipitation over the conterminous United States also show that the multimodel mean represents quite well the observed behavior of lagged autocorrelations of pentad rainfall. Persistence of categorical anomalies in monthly rainfall during boreal summer across Northern Hemisphere midlatitudes is also well represented by the multimodel mean. There is again a large degree of variation among models in the strength of these metrics for precipitation memory, but the results of K03 and KS04 suggest that the land surface is a likely culprit in supplying this persistence to the precipitation signal.

Overall, it appears that there is still much that can be done to improve the behavior (i.e., the parameterizations) related to land-atmosphere interactions in the GCMs widely used for weather and climate prediction and research. Variations among models can arise for many, often subtle reasons having to do with details of the parameterizations and the interplay of components and tunings of the models (Teuling and Troch 2005). Liu et al. (2005) have shown what can be accomplished toward improved model performance simply by considering the land and atmosphere parameters together in existing parameterizations when calibrating coupled models. The multimodel approach like that of GLACE is not an antidote but does alleviate the symptoms of individual model errors and biases.

We cannot disprove the results of GLACE over the limited areas where there are sufficient data to estimate locally the land-atmosphere coupling strength. Rather, we can argue that we still do not have sufficient data to quantify the actual strength of coupling between land and atmosphere. Long-term collocated measurements of SW_{et}, surface fluxes, and near-surface meteorology should be distributed around the globe in order to aid model development and assess the potential for SW_{et}

as a predictor for climate via land-atmosphere feedback. In the mean time, land-atmosphere model development efforts could benefit by paying more attention to local validation of land surface and boundary layer parameterizations with available in situ data.

Acknowledgments. The authors thank G. Hughes for providing us with the ARM data for the SGP Extended Facility sites. We would also like to thank Tim DelSole for useful and enlightening discussions as this work progressed. The time and effort of individuals at all of the participating GLACE modeling centers have made this study possible: G. Bonan, E. Chan, P. Cox, C. T. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence, P. Liu, C.-H. Lu, S. Malyshev, B. McAvaney, J. L. McGregor, K. Mitchell, D. Mocko, T. Oki, K. W. Oleson, A. Pitman, Y. C. Sud, C. M. Taylor, D. Verseghy, R. Vasic, Y. Xue, and T. Yamada. This work was conducted under support from National Aeronautics and Space Administration Grant NAG5-11579.

REFERENCES

- Ackerman, T. P., and G. M. Stokes, 2003: The Atmospheric Radiation Measurement Program. *Phys. Today*, **56**, 38–44.
- Baldocchi, D., and Coauthors, 2001: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Amer. Meteor. Soc.*, **82**, 2415–2434.
- Betts, A. K., 2004: Understanding hydrometeorology using global models. *Bull. Amer. Meteor. Soc.*, **85**, 1673–1688.
- , J. H. Ball, A. C. M. Beljaars, M. J. Miller, and P. A. Viterbo, 1996: The land surface-atmosphere interaction: A review based on observational and global modeling perspectives. *J. Geophys. Res.*, **101**, 7209–7226.
- Bolton, D., 1980: The computation of equivalent potential temperature. *Mon. Wea. Rev.*, **108**, 1046–1053.
- Cook, D. R., 2005: Energy Balance Bowen Ratio (EBBR) handbook. ARM Tech. Rep. TR-037, 23 pp.
- Crow, W. T., and E. F. Wood, 2002: Impact of soil moisture aggregation on surface energy flux prediction during SGP'97. *Geophys. Res. Lett.*, **29**, 1008, doi:10.1029/2001GL013796.
- D'Odorico, P., and A. Porporato, 2004: Preferential states in soil moisture and climate dynamics. *Proc. Natl. Acad. Sci. USA*, **101**, 8848–8851.
- Falge, E., and Coauthors, 2001: Gap filling strategies for long-term energy flux data sets. *Agric. For. Meteorol.*, **107**, 71–77.
- Fennessy, M. J., and Y. Xue, 1997: Impact of USGS vegetation map on GCM simulations over the United States. *Ecol. Appl.*, **7**, 22–33.
- Findell, K. L., and E. A. B. Eltahir, 1997: An analysis of the soil moisture-rainfall feedback, based on direct observations from Illinois. *Water Resour. Res.*, **33**, 725–735.
- , and —, 2003a: Atmospheric controls on soil moisture-boundary layer interactions. Part I: Framework development. *J. Hydrometeorol.*, **4**, 552–569.
- , and —, 2003b: Atmospheric controls on soil moisture-boundary layer interactions. Part II: Feedbacks within the continental United States. *J. Hydrometeorol.*, **4**, 570–583.

- GFDL Global Atmospheric Model Development Team, 2004: The new GFDL global atmosphere and land model AM2-LM2: Evaluation with prescribed SST simulations. *J. Climate*, **17**, 4641–4673.
- Guo, Z., and Coauthors, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part II: Analysis. *J. Hydrometeorol.*, **7**, 611–625.
- Higgins, R. W., W. Shi, E. Yarosh, and R. Joyce, 2000: *Improved United States Precipitation Quality Control System and Analysis*. NCEP/Climate Prediction Center Atlas 7, U.S. Department of Commerce, 40 pp.
- Huffman, G. J., and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc.*, **78**, 5–20.
- Kochendorfer, J. P., and J. A. Ramirez, 2005: The impact of land–atmosphere interactions on the temporal variability of soil moisture at the regional scale. *J. Hydrometeorol.*, **6**, 53–67.
- Koster, R. D., and M. J. Suarez, 2003: Impact of land surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeorol.*, **4**, 408–423.
- , and —, 2004: Suggestions in the observational record of land–atmosphere feedback operating at seasonal time scales. *J. Hydrometeorol.*, **5**, 567–572.
- , —, R. W. Higgins, and H. Van den Dool, 2003: Observational evidence that soil moisture variations affect precipitation. *Geophys. Res. Lett.*, **30**, 1241, doi:10.1029/2002GL016571.
- , and Coauthors, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140.
- , and Coauthors, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. *J. Hydrometeorol.*, **7**, 590–610.
- Liu, Y., H. V. Gupta, S. Sorooshian, L. A. Bastidas, and W. J. Shuttleworth, 2005: Constraining land surface and atmospheric parameters of a locally coupled model using observational data. *J. Hydrometeorol.*, **6**, 156–172.
- Meyers, T. P., 2001: A comparison of summertime water and CO₂ fluxes over rangeland for well watered and drought conditions. *Agric. For. Meteorol.*, **106**, 205–214.
- , and S. E. Hollinger, 2004: An assessment of storage terms in the surface energy balance of maize and soybean. *Agric. For. Meteorol.*, **125**, 105–115.
- Salvucci, G. D., J. A. Saleem, and R. Kaufmann, 2002: Investigating soil moisture feedbacks on precipitation with tests of Granger causality. *Adv. Water Res.*, **25**, 1305–1312.
- Sud, Y. C., D. M. Mocko, K.-M. Lau, and R. Atlas, 2003: Simulating the Midwestern U.S. drought of 1988 with a GCM. *J. Climate*, **16**, 3946–3965.
- Teuling, A. J., and P. A. Troch, 2005: Improved understanding of soil moisture variability dynamics. *Geophys. Res. Lett.*, **32**, L05404, doi:10.1029/2004GL021935.
- , R. Uijlenhoet, and P. A. Troch, 2005: On bimodality in warm season soil moisture observations. *Geophys. Res. Lett.*, **32**, L13402, doi:10.1029/2005GL023223.